1 Precursors

What Is Trigonometry?

This deceptively difficult question will shape the opening chapter. The notion that sines, cosines, and other modern functions define what we mean by "trigonometry" may be laid to rest instantly; these functions did not reach their modern forms, as ratios of sides in right-angled triangles, until relatively recently. Even in their historical forms they did not appear until medieval India; the Greeks used the length of the chord of an arc of a circle as their only trigonometric function.[1] The word itself, meaning "triangle measurement," provides little help: it is a sixteenth-century term, and much ancient and medieval trigonometry used circles and their arcs rather than triangles as their reference figures.

If one were to define trigonometry as a science, two necessary conditions would arise immediately:

- a standard quantitative measure of the inclination of one line to another;[2] and
- the capacity for, and interest in, calculating the lengths of line segments.

We shall encounter sciences existing in the absence of one or the other of these; for instance, pyramid slope measurements from the Egyptian Rhind papyrus fail the first condition, while trigonometric propositions demonstrated in Euclid's *Elements* (the Pythagorean Theorem, the Law of Cosines) fail the second.

What made trigonometry a discipline in its own right was the *systematic* ability to convert back and forth between measures of angles[3] and of lengths. Occasional computations of such conversions might be signs of something better to come, but what really made trigonometry a new entity was the ability to take a given value of an angle and determine a corresponding length. Hipparchus's work with chords in a circle is the first genuine instance of this, and we shall begin with him in chapter 2. However, episodes that come

[1] The chord is the length of a segment cut off across a circle, usually considered as a function of the angle it subtends at the center.

[2] Today this is done with angles, although in certain contexts such as spherical trigonometry, circular arcs (treating the given lines as radii of the circle) are more convenient.

[3] For lack of a more convenient term, we shall use "angular measurement" to refer to any direct measure of inclination, whether angles or arcs are measured.

close—a prehistory—do exist in various forms before Hipparchus, and we shall mention some of them in this preliminary chapter.

There is a case to be made that definitions are unwise in a historical account, especially one covering several cultures and a vast time span. Any idea that produces fruit for as long a time as trigonometry is likely to be altered to the point of unrecognizability after a while—consider, for instance, the tortured history of the mathematical meaning of "analysis"—and trigonometry is no exception. It is hard to reconcile the use of sines and cosines in the analytical work of men like Euler, Fourier and Cantor with the geometric spirit of trigonometry's Greek progenitors. So, a definition that fits one part of the history may well be cumbersome elsewhere. Fortunately, our definition coincides (mostly) with the realization and growth of a powerful single idea, but in cases where they diverge, we will stay with the flow rather than the definition.

The *Seqed* in Ancient Egypt

The pyramids are among the most recognized structures in the world; the Great Pyramid at Giza is the oldest of the seven ancient wonders of the world, and the only one still standing. They have become the center of speculative theories of all sorts, purportedly revealing advanced age, advanced religious significance, advanced astronomy, advanced mathematics, and advanced engineering. Many of these theories have been shown to be little more than advanced silliness; nevertheless, the pyramids are undoubtedly some of the most remarkable buildings ever constructed. The precision in both their North-South alignments and their slopes on such a staggering scale are an achievement regardless of the level of sophistication required to produce them.

Part of the reason for the vast literature on the construction of the pyramids is, no doubt, the paucity of real evidence. In fact, much of what we know about Egyptian mathematics in general comes from only a few papyri that somehow survived the millennia. The most famous of these texts is the Rhind Mathematical Papyrus (RMP),[4] named after the nineteenth-century Scotsman A. Henry Rhind, who obtained it in Luxor in 1858. This document, over 5 m wide and about 32 cm high, consists mostly of the famous

[4] Much modern scholarship on the RMP is based on the edition [Peet 1923] and the two-volume edition, translation, and commentary by A. B. Chace [Chace 1927/1929]. A new translation has recently appeared within volume 3 of Marshall Clagett's *Ancient Egyptian Science: A Source Book* [Clagett 1999]. [Gillings 1972] contains substantial commentaries on the RMP. See also [Robins and Shute 1987] for photographic plates of the entire papyrus and an explanation of its mathematics.

$2/n$ table used to convert fractional quantities into sums of unit fractions, a table of division of integers by 10, and a set of 84 mathematical problems. The problems are mostly arithmetical or involve the distribution of food, but some deal with the geometry of circles, rectangles, triangles, and pyramids. Problems 56–60 in particular contain calculations regarding the slope of a pyramid given its horizontal and vertical dimensions, and some have posited these problems as a kind of proto-trigonometry.[5]

The cause of these claims is the notion of the *seqed*, a term referring to the slope of an inclined side in Egyptian architecture. Used in the RMP only with respect to pyramids, there is evidence to suggest that the *seqed* was also used for the inclinations of temple gateways.[6] As a measure of "slope" it inverts our use of the word: the *seqed* is the amount of horizontal displacement, measured in palms, for every seven palms of vertical displacement (thus $s = 7/m$, where m is the modern definition of slope). Seven palms corresponded to one architectural royal cubit. Curiously, analysis of inclinations in Egyptian art suggests that something similar to the *seqed* may have been at work in that field as well, but using the small cubit of six palms instead. In the case of the temple of Ramesses II at Luxor, the use of a different cubit leads to a *seqed* of 1 (per 7) for the gateway itself, but an inclination of 1 (per 6) in a relief of the same gateway on a wall inside the temple!

Text 1.1
Finding the Slope of a Pyramid
(Rhind Mathematical Papyrus, Problem 58)

In a pyramid whose altitude is 93⅓, make known the seqed of it when its base-side is 140 [cubits].

Take ½ of 140, which is 70. Multiply 93⅓ so as to get 70. ½ of 93⅓ is 46⅔. ¼ of it is 23⅓. Take ½ ¼ of a cubit. Operate on 7: ½ of it is 3½; ¼ of it is 1½ ¼; the total is 5 palms 1 finger. This is the seqed.

[5] From [Robins and Shute 1985]: "We know from the pyramid exercises in the Rhind mathematical papyrus . . . that the ancient Egyptians used a simple trigonometry for determining architectural inclinations . . ." See also [Vetter 1925] and [Vogel 1959, 72–73]. The image on the cover of the recent *Trigonometric Delights* by Eli Maor ([Maor 1998]) contains both the Sphinx and the Great Pyramid.

[6] See [Robins and Shute 1985, 113]. For the use of the notion of slope in water clocks, see [Clagett 1995, 76]. For a description of how the *seqed* would have been implemented in pyramid construction, see [Rossi 2004, 192–196]. The extent to which the *seqed* was actually used is controversial; see the chapter on *seqed* theory in [Herz-Fischler 2000, 30–45] for a careful analysis, including a summary of various nineteenth-century opinions. Finally, [Imhausen 2003a, 162–168] contains an algorithmic analysis of the RMP *seqed* calculations.

Working out:

1 93⅓
\½ 46⅔
\¼ 23⅓

Total ½ ¼
Produce ½ ¼ of a cubit, the cubit being 7 palms.

1 7
½ 3½
¼ 1[½]¼

Total: 5 palms 1 finger, which is the seqed.[7]

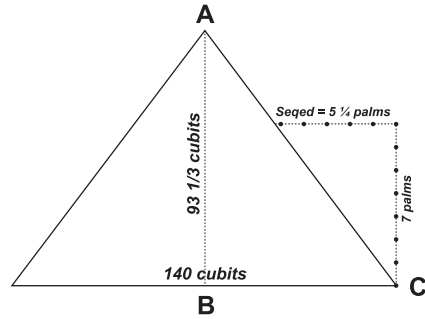
Explanation: In figure 1.1, the base is bisected to give $BC = 70$ cubits. The calculations that follow establish that $BC / AB = 70 / (93⅓) = ¾$. To convert this ratio to a *seqed* measurement, multiply this result by 7, which gives a value of 5¼ (commonly used in pyramids of the late Old Kingdom).

The trigonometric connection resides in the fact that the *seqed* is simply the ratio BC / AB (scaled by the unit conversion from palms to cubits), and is thus equal to the cotangent of the angle at C. Is this legitimate trigonometry? The text does not refer explicitly to $\angle C$, but as we shall see, Hellenistic astronomers often used arcs rather than angles for computations that are clearly trigonometric. Here, however, there is no notion of the measurement of $\angle C$ in any form. The purpose of the *seqed* is architectural, and to read it as a cotangent is anachronistic. For this reason, and also since the *seqed* did not transmit to the Greeks, we shall say no more about it here.[8]

Babylonian Astronomy, Arc Measurement, and the 360° Circle

Both angles and lengths are continuous quantities; to be able to move easily between them requires a good system of numeration for fractional quantities. The ancient Egyptians used unit fractions exclusively, and forbade repeated use of the same unit fraction when representing a number. Clearly this would have made converting arbitrary angles to lengths cumbersome. The first powerful number system incorporating fractions arose in the third

[7] This translation is taken from [Clagett 1999, 167].
[8] Similar concepts of slope are found in other ancient cultures; for instance, see [Robson 1999, 222] for examples from Old Babylon. We shall ignore them here for the same reason that we pass over the *seqed*.

Figure 1.1
RMP Problem 58, illustrating the *seqed*

millennium BC in Sumer, out of a system of metrological units, particularly measurements of weight.[9] A quantity of ten of one unit became the next unit up, then six of that unit became the next, and so on. Eventually pairs of values in the alternating bases of 6 and 10 came to be thought of as single values in a number system with a base of 60. Since 60 has many divisors, sexagesimal calculation (as it is now called) is not as difficult as it may sound.[10] An administrative reform in the twenty-first century BC, requiring the keeping of meticulous records, may have instigated its birth.[11] Whatever the cause, it was to become the standard system of numeration in astronomy and trigonometry, at least for the fractional parts of numbers, for millennia.[12]

We shall represent sexagesimal numbers in the standard modern transcription; that is,

$$23;51,20 = 23 + \frac{51}{60} + \frac{20}{60^2}.$$

Since sexagesimals had arisen from metrology there was nothing particularly special about any given unit; the measurement could always be subdivided into something smaller. Hence the sexagesimal point, here indicated by a semicolon, was not as significant as the decimal point is to us, and was not written. The lack of a symbol for zero would also have caused some confusion until placeholder symbols for the absence of a quantity in a sexagesimal digit were invented.

[9] See [Thureau-Dangin 1921], [Neugebauer 1927], [Thureau-Dangin 1928a], and the nice survey article [Thureau-Dangin 1939].

[10] It is often said erroneously in textbooks that the base of 60 was chosen specifically because of its many divisors.

[11] [Høyrup 1992, 604]. The case for this dating is made in [Powell 1976].

[12] Indeed, [Nissen/Damerow/Englund 1993, 151] asserts that sexagesimal arithmetic was used initially only in arithmetic texts, and if not for the birth of astronomy, it may have “fallen into oblivion.”

The sexagesimal system is still seen today, for instance, in the division of the hour into minutes and the minute into seconds, and of angular degrees into minutes and seconds. Place value numeration was helpful then, as now, for complicated arithmetic computations; the Babylonians were adept at addition, subtraction, multiplication, and division (although certain complications arise for division[13]). Tables of square and cube roots reveal that the Babylonians also had methods for these operations.[14]

By far the most famous computational tablet of the early Babylonian period is Plimpton 322, composed around 1700 BC. Excavated during the 1920s, it contains a collection of Pythagorean triples—that is, sets of three integers that satisfy the Pythagorean theorem $a^2 + b^2 = c^2$, thereby forming a right triangle. The smallest examples of such triples are 3-4-5 (a version of which appears on the tablet) and 5-12-13. One way to interpret this tablet is by trigonometry: after all, the triples appear to compare lengths of sides in a triangle, and they are arranged within the table in order of decreasing ratio of the two sides touching the right angle. However, there are a host of competing explanations, including an interpretation in terms of reciprocal pairs.[15] With no explicit system of angle measurement in this tablet or indeed in any early Babylonian mathematics,[16] nor any tradition of similar tablets of triples to follow it, Plimpton 322 fits neither our definition of trigonometry nor our narrative; hence we move on.[17]

However, angular measurement did begin with the Babylonians; but it was over a millennium later and in the context of astronomy. Observations of celestial phenomena date back almost as far as human history, but the earliest documents that record these events systematically and develop a quantitative science to accompany them began around the eighth century BC.[18]

[13] The Babylonian procedure for division is simply to find the reciprocal of the divisor and multiply by the numerator. Tables of reciprocals existed to aid the process, but usually gave only reciprocals whose sexagesimal representation terminated after a finite number of places. See [Aaboe 1964, 10–16 and 20–22], among many other sources, for more details.

[14] This is well known; indeed, [Neugebauer/Sachs 1945, 33–35] lists some tables, and p. 42 analyzes a procedure text for cube roots. A consideration of an algorithm that may have been used for evaluating square roots may be found in [Fowler/Robson 1998].

[15] The literature on Plimpton 322 is extensive, perhaps well beyond its significance for Babylonian mathematics. See [Neugebauer/Sachs 1945, 38–41] for its original publication. Most recently, see [Robson 2001] and [Robson 2002] for a revival of the reciprocal pairs theory; the former contains descriptions of competing theories and an extensive bibliography.

[16] [Robson 2001, 179–183] describes the Old Babylonian approach to circles as conceptualized "from the outside in"; even the radius was a subsidiary idea. See also [Gandz 1929, 458–459] for a refutation of the notion of angle in either Babylonian or Egyptian mathematics.

[17] The use of Pythagorean triples is also controversial in Egyptian architecture and mathematics. See [Rossi 2004, 60–65 and 216–221] for a good summary of the evidence.

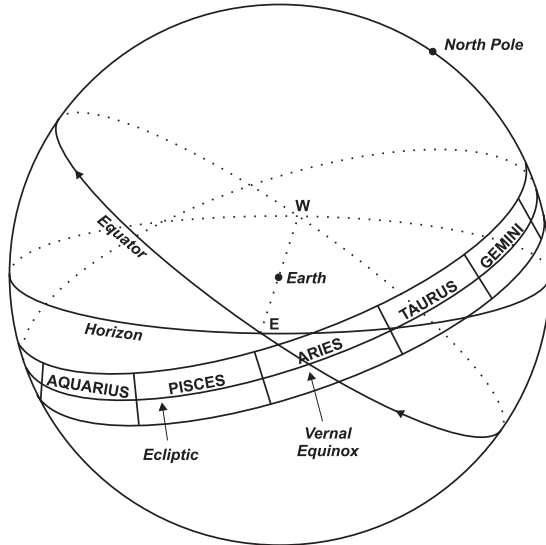
[18] See [Neugebauer 1975, part I, pp. 351–353] for a short summary of our sources for Babylonian astronomy.

The driving forces behind these endeavors would have been partly calendric (for civil and agricultural purposes), but astrological needs were dominant. Perhaps as an outgrowth of this horoscopic heritage Babylonian astronomy was computational and predictive, rather than geometrical and explanatory. Rather than predict locations of celestial bodies at given times, astronomers predicted the times and places of given celestial events. There was no underlying geometric theory of how the heavens should work, only computational schemes to say when and where a periodic event would next occur.

Nevertheless, being able to answer the "where" question requires a means of pointing to a particular place in the sky, and to grasp the Babylonian solution we shall need some basic astronomy—although we should be clear that the Babylonians did not necessarily share our need for the geometric pictures to follow.[19] The dome of the sky is called the ***celestial sphere*** (see figure 1.2). Over the course of a day the celestial sphere performs a single rotation around us from east to west, carrying the stars and the Sun with it, thereby causing day and night. The stars remain in the same place with respect to each other, hence the name ***fixed stars***. To any observer the ***celestial equator*** remains in the same place in the sky, turning into itself with its daily rotation. Seven bodies (the Sun, Moon, and the five visible planets) also participate in this motion, but they also move slowly with respect to the fixed stars. For instance, the Sun travels on its own great circle in the celestial sphere, known as the ***ecliptic***, completing one revolution per year in the direction opposite to that of the celestial sphere's daily motion, or roughly 1° per day. The ecliptic does not remain fixed in place; it is carried by the daily rotation of the celestial sphere. Twice per year the Sun crosses the celestial equator, and at these times day and night are of the same length; these points are called the ***vernal and autumnal equinoxes***. (A fuller description of these astronomical concepts may be found in the chapter on the ancient heavens.)

For the simple reason that the shape of the solar system is close to a disc, most interesting celestial events—those involving the seven moving bodies—take place within a few degrees of the ecliptic. This band, known as the ***zodiac*** (see figure 1.2), was therefore the astronomer's arena. The MUL.APIN, a well-known text from the seventh century BC, contains an early version of the zodiac before the appearance of the twelve constellations that became the zodiacal signs, mentioning these twelve and another six constellations

[19] There is a debate over the extent to which the Babylonians visualized their astronomy; even the celestial sphere may not have been necessary. See [Brack-Bernsen 2003, esp. 23–24].

Figure 1.2
The celestial sphere, the equator, the ecliptic, and the zodiac

"who stand in the path of the Moon."[20] The Sun's motion was described as follows:

> From XII 1 to II 30 the Sun is in the path of Anu:
> Wind and storm.
> From III 1 to V 30 the Sun is in the path of Enlil:
> Harvest and heat.
> From VI 1 to VIII 30 the Sun is in the path of Anu:
> Wind and storm.
> From IX 1 to XI 30 the Sun is in the path of Ea:
> Cold.[21]

Hence the Sun moves on an inclined circle divided into four regions, spending three months in each.[22] From here it would be an easy step to split each region into three parts, one corresponding to each month; this division happened by the early fifth century BC. Each of the twelve regions was identified

[20] [Van der Waerden 1974, 79–80]. *Science Awakening II: The Birth of Astronomy* gives a fairly thorough account of the emergence of the zodiac in scattered places, and is partly the basis of the following discussion. See also [Evans 1998, 5–8] for a description of MUL.APIN, and [van der Waerden 1953] and [Brack-Bernsen 2003] on the development of the zodiac and ecliptic in Babylon and Greece.
[21] Quoted from [van der Waerden 1974, 80].
[22] "Month" refers to the time it takes for the moon to orbit the Earth.

with a constellation on or near the zodiac, giving rise to the ***zodiacal signs*** Aries, Taurus, and so forth.[23]

The signs, eventually made equal in size, were each further subdivided into 30 *uš* (meaning length). This division is convenient for a couple of reasons: the Sun travels roughly one uš per day, and 30 is a handy number in sexagesimal calculations. In fact the uš was used only for computation; for recording observations, another measurement called the *kùš* (equal to about 2.4 uš) was standard.[24] Nevertheless the zodiac was now divided into 360 units, later to be called degrees, and a unified system of arc measurement was in place,[25] at least for the zodiac.

Now, Babylonian astronomy was not designed with underlying geometric models in mind, and while it might seem natural to us to consider a planet's path as a geometric object in the heavens and work toward astronomical predictions from that, this was not necessarily so for the Babylonians. Consider, for instance, their approach to the motion of the Sun. It travels about 1° per day in its annual trip around the ecliptic, but its speed is not constant; in ancient times it was slower in the spring, and faster in the fall.[26] The Babylonians used two methods to represent the Sun's changing speed: System A, which alternates the speed between two fixed values; and System B, which varies the speed linearly between a maximum and a minimum (see figure 1.3).[27] It may be tempting, but is anachronistic, to consider these methods as approximations to trigonometric curves;[28] functions and graphs are modern inventions. In any case, these systems were numeric constructs,

[23] This zodiac differed from the Greek one, in that the vernal equinox happened to fall either 8° or 10° into Aries, and was fixed with respect to the background stars. (For research on this topic, see for instance [van der Waerden 1953] and [Huber 1958].) This becomes important with the later Greek discovery of the ***precession of the equinoxes***, a very slow movement of the vernal equinox with respect to the background stars (1° every 72 years). The Greeks fixed the beginning of Aries to the vernal equinox rather than the stars. Thus, today the zodiacal signs have moved about 30° away from the constellations for which they were named.

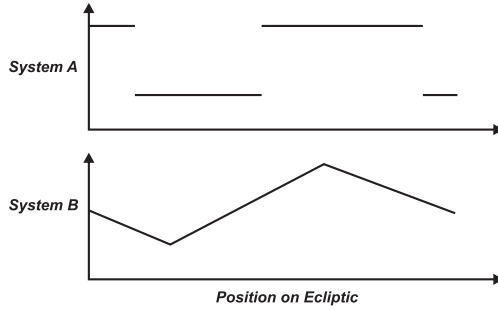
[24] See [Swerdlow 1998, 34–37].

[25] For more detail, including the relation to time measurement, see [Thureau-Dangin 1928b], [Sidersky 1929], [Thureau-Dangin 1930], [Thureau-Dangin 1931], [Neugebauer 1938], and [Neugebauer 1983, 8]. A candidate for the earliest text with a 360° zodiac is one published in [Sachs 1952, 54–57], probably dated 410 BC. Alongside ecliptic coordinates, another system was used for certain types of texts ("Diaries," "Goal-Year texts," and "Normal-Star Almanacs,") whereby locations were identified with respect to 31 "normal stars" scattered irregularly around the zodiac. See [Neugebauer 1975, 545–547].

[26] In modern times the summer is the longest season in the northern hemisphere.

[27] Fuller descriptions of Babylonian models of the motions of the planets may be found in any number of places; for instance, a survey in [Neugebauer 1957/1969, 97–144] and thorough coverage in [Neugebauer 1975, 347–540] and [Brown 2000], and conjectures on the theory's construction in [Aaboe 1965], [Aaboe 1980], and [Swerdlow 1998].

[28] It has been tried, for example, in [Dittrich 1934].

Figure 1.3
Two Babylonian models for the Sun's velocity throughout the year

allowing the direct prediction of solar positions without the intervening presence of a geometric model.

The Geometric Heavens: Spherics in Ancient Greece

Meanwhile, Greek astronomy took a very different turn. The earliest astronomical interests in Greece were likely similar to other cultures: firstly calendrical, concerning the coordination of dates and weather events with the risings and settings of certain stars and constellations. The belief that the heavens and the Earth are concentric spheres is very old; the bringing together of this cosmological statement with these astronomical problems may have come with Eudoxus of Cnidus in the mid-fourth century BC.[29] The next decades produced several texts on the mathematics of spheres with an eye to their use in the heavens. Noteworthy among these were Autolycus of Pitane's *On a Moving Sphere* and *On Risings and Settings*,[30] and Euclid's *Phaenomena*.[31] These works were among the treatises called the "Little Astronomy," used by later writers as an approach to the study of Ptolemy's *Almagest*.[32]

[29] [Goldstein/Bowen 1983] argues that Eudoxus's two-sphere model marks the beginning of a new phase of Greek astronomy, linking calendar problems with cosmology and allowing mathematical reasoning to operate on both. One scholar believes that Eudoxus performed trigonometric calculations to determine terrestrial latitudes from the lengths of shadows cast by the gnomon (a simple astronomical instrument to be discussed in the next chapter); see [Szabó 1985a].
[30] The Greek text and an English translation of these works may be found in [Autolycus (Bruin/Vondjidis) 1971]; Greek text and French translation are in [Autolycus (Mogenet) 1950]; and German translations of Autolycus and Theodosius are in [Czwalina 1931].
[31] See [Berggren/Thomas 1996] for a translation and study of Euclid's *Phaenomena*.
[32] [Neugebauer 1975, 768–769] doubts that the "Little Astronomy" actually existed.

The problems dealt with in these treatises had much to do with the earlier astronomical and geographical concerns;[33] for instance, the varying lengths of daylight throughout the year for a given location, the division of the Earth's surface into climate zones, or the angles between important great circles in the celestial sphere. Euclid's *Phaenomena* is geared particularly to the times required for various arcs of the ecliptic to rise above the horizon. Since the Sun is a point on the ecliptic, rising times were crucial for the reckoning of the passage of time throughout the day, and we shall return to this topic several times. One typical theorem reads as follows:

> **[Proposition 9b]**. Semicircles of the ecliptic that do not begin on the same parallel rise entirely in unequal times, [with] that following Cancer in the greatest, those following, each in turn, in lesser [times], and that following Capricorn in the least. And those [semicircles] beginning on the same parallel rise in equal times.[34]

To see what this is saying consider figure 1.4, which positions the celestial sphere so that the equinoxes (the places where the equator and ecliptic intersect) are due north and south. The semicircle "following Cancer" is drawn in bold;[35] dashes or dots are on the back surface of the sphere from our perspective. The equator rotates at a constant speed, carrying the ecliptic along with it. The point on the equator that is now at E on the horizon will take twelve hours to rise above the horizon, travel through the sky, and reach the horizon again at W. On the other hand the bolded part of the ecliptic has already begun to rise above the horizon, and twelve hours from now it will have as far to go before setting as it has risen in the figure. So the bolded semicircle takes more than twelve hours to rise. Conversely, the part of the ecliptic not drawn in bold will take less than twelve hours to rise.

Euclid's proof of this proposition is entirely geometric. Even so, it has been suggested that he may have been aware of Babylonian arithmetic schemes dealing with length of daylight (similar to the solar schemes of figure 1.3), and he may have been trying to put the symmetries used in these schemes on firm geometrical ground.[36]

[33] See [Aujac 1976] for a discussion of the interaction between astronomy and geography, and the latter's use of both spherics and "spheropoeia" (which included the use of instruments).

[34] [Berggren/Thomas 1996, 71].

[35] By now the signs of the ecliptic were positioned so that Aries began at the vernal equinox, as opposed to the Babylonian practice of figure 1.2.

[36] [Berggren/Thomas 1996, 2] and [Berggren 1991b, especially 237–238].

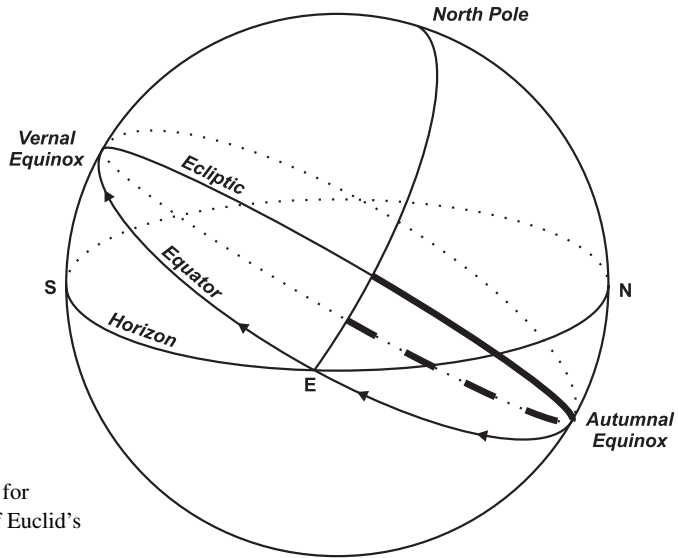



Figure 1.4
Simplified figure for Proposition 9b of Euclid's *Phaenomena*

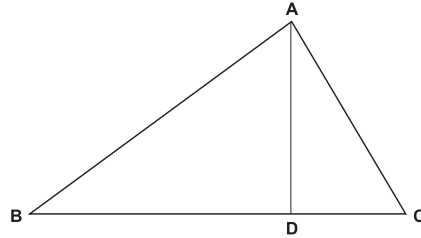
A Trigonometry of Small Angles? Aristarchus and Archimedes on Astronomical Dimensions

Although early spherics was able to say a great deal about the qualitative behavior of the celestial sphere, it was not the basis for a predictive science.[37] For instance, one could not use it to infer precisely where a particular body would be at a given time. In this case the geometers would have been of little help. Our notion of geometry, using coordinates, leads us to think easily of lengths as quantities, measured by numbers; and locations as pairs (or triples) of numbers referred to coordinate axes. Indeed, we often think of a curve as an algebraic equation: for instance, a parabola *is* an equation of the form $y = ax^2 + bx + c$. By contrast, Greek geometry operated on its own and had no need for such a numeric foundation.[38]

Indeed, if we were to impose a numerical understanding of magnitudes on Greek geometry, then we might say that many trigonometric results were

[37] See [Knorr 1989] for a discussion of four Greek scientific disciplines that do not employ numeric techniques, including early spherics. Knorr argues that their purpose was to explain, not predict, phenomena.

[38] The case is made in detail in [Fowler 1987] that Greek mathematics and astronomy, up to and including Archimedes, were not "arithmetized"; see also [Fowler 1992].

Figure 1.5
Diagram for *Elements* II.13

known well before the birth of trigonometry. The most obvious is the Pythagorean Theorem, which (for a unit circle) is equivalent to the identity $\sin^2\theta + \cos^2\theta = 1$. More dramatically, Proposition II.13 of Euclid's *Elements* reads:

> In acute-angled triangles the square on the side subtending the acute angle is less than the squares on the sides containing the acute angle by twice the rectangle contained by one of the sides about the acute angle, namely that on which the perpendicular falls, and the straight line cut off within by the perpendicular towards the acute angle.[39]

In figure 1.5, if $a = BC$, $b = AC$, and $c = AB$, then $BD = c \cos\angle B$, and we have the Law of Cosines:

$$b^2 = a^2 + c^2 - 2ac \cos\angle B.$$

But these sorts of statements are not much closer to trigonometry than was the Egyptian *seqed*. We have a working definition of angle, but no quantitative measurement or systematic conversion to lengths.[40]

The first real glimmer of a proper trigonometry came with the earliest stages of a quantitative astronomy in the third century BC. As so often in the history of science, it is very difficult to assert a precise beginning; a blurry gradual awakening does not lend itself to the identification of a single blessed moment. The inevitable result is that the birth of mathematical astronomy has been attributed to several different people.[41] One of the earliest figures in this

[39] [Euclid 1925, 406].
[40] See Chapter 4, on John of Muris, for an example of II.13 being used more legitimately as the Law of Cosines to solve a triangle.
[41] [Neugebauer 1959], for instance, argues that "all the evidence points to Apollonius as the founder of Greek mathematical astronomy" based on his work on planetary motion. [Goldstein/Bowen 1983], on the other hand, claim that Eudoxus was "[t]he person largely responsible for turning astronomy into a mathematical science."

story is Aristarchus of Samos (ca. 310 BC–230 BC),[42] who is most famous for being the first to propose a Sun-centered solar system, about 1,700 years before Copernicus. The little we know of "the mathematician," as he was called, includes that he was a student of the philosopher Strato of Lampsacus; Vitruvius credits him with the invention of the scaphe, a type of sundial.[43] Unfortunately most of his works are lost, including what he wrote on his heliocentric system, and a book on optics. Our knowledge of Aristarchus's solar system therefore must be rebuilt from references in writings of later authors, especially Archimedes' *Sand Reckoner*, of which we shall speak shortly.[44]

Aristarchus's only surviving work, *On the Sizes and Distances of the Sun and Moon* (see figure 1.6), has nothing to do with his heliocentric theory; it sets out to measure the relative and absolute sizes of the Earth, Moon, and Sun.[45] It was one of the treatises in the "Little Astronomy," and it is probably due to its inclusion in this collection that we have it today.[46] It reads in the classical Greek style, reminiscent of Euclid, and its theorems are worthy of one known as "the mathematician."

The key datum for Aristarchus's determination of the relative distances of the Moon and the Sun comes from the moment when precisely half of the Moon's face is visible from the Earth. In figure 1.7, when the observer sees the half Moon, the Sun–Moon–Earth angle is a right angle; the Moon–Earth–Sun angle is asserted to be "less than a quadrant by one-thirtieth of a quadrant" $\approx 3°$ i.e., $87°$.[47] (The correct value is close to $89°50'$.) This hopelessly poor value, in addition to some other incorrect observations and assumptions, lead one to question whether Aristarchus's purpose was less quantitative than proof of concept, "a purely mathematical exercise"[48]—albeit an ingenious one.

To convert his angular data to ratios of distances, Aristarchus is forced into approximations of trigonometric quantities. As a sample of his technique we shall follow the first half of Proposition 7, in which he demonstrates that

[42] The source on which most twentieth-century work on Aristarchus is based is [Heath 1913], which contains a translation of *On the Sizes and Distances of the Sun and Moon*. See also the recent detailed textual and mathematical discussion of *On the Sizes* in [Berggren/Sidoli 2007]. For a biographical essay, see [Stahl 1970].

[43] [Vitruvius 1914, 273] or [Vitruvius 1934, vol. 2, 255].

[44] Aristarchus's heliocentric hypothesis has attracted a fair bit of scholarly attention, especially the extent to which Copernicus may have been indebted to him. Among others, see [von Erhardt/von Erhardt-Siebold 1942 and 1943], [Neugebauer 1942], [Africa 1961], [Derenzini 1974], [Wall 1975], [Rosen 1978], [Gingerich 1985], [Zhitomirskiĭ 1986], and [Christianides et al. 2002]. [Wall 1975] contains a summary of all extant information on Aristarchus.

[45] [Tannery 1883, 243–244] argues that Eudoxus may have originated the method that follows.

[46] [Heath 1913, 317–318].

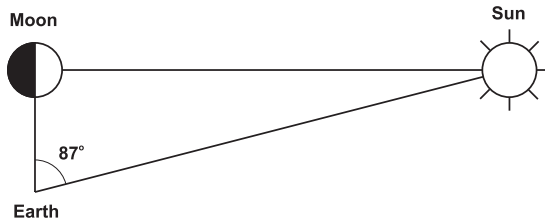
[47] [Heath 1913, 353].

[48] Neugebauer makes this case in [Neugebauer 1975, 634–643], especially p. 643.

ARISTARCHI
DE MAGNITVDINIBVS,
ET DISTANTIIS SOLIS,
ET LVNAE, LIBER
CVM PAPPI ALEXANDRINI
explicationibus quibuſdam.
A' FEDERICO COMMANDINO
Vrbinate in latinum conuerſus, ac
commentarijs illuſtratus.
Cum Priuilegio Pont. Max. In annos X.

PISAVRI, Apud Camillum Franciſchinum.
M D LXXII.

Figure 1.6
Title page of Commandino's 1572 edition of Aristarchus of Samos's *On the Sizes* (courtesy of the Burndy Library)

Figure 1.7
Earth, Sun, and Moon at half moon, from Aristarchus's *On the Sizes*

the ratio of the Sun's distance to the Moon's distance, is greater than 18 but less than 20.

Text 1.2
Aristarchus, the Ratio of the Distances of the Sun and Moon
(Proposition 7a from *On the Sizes*)

[See figure 1.8.] *The distance of the Sun from the Earth is greater than eighteen times, but less than twenty times, the distance of the Moon from the Earth.*

For let *A* be the center of the Sun, *B* that of the Earth.

Let *AB* be joined and produced.

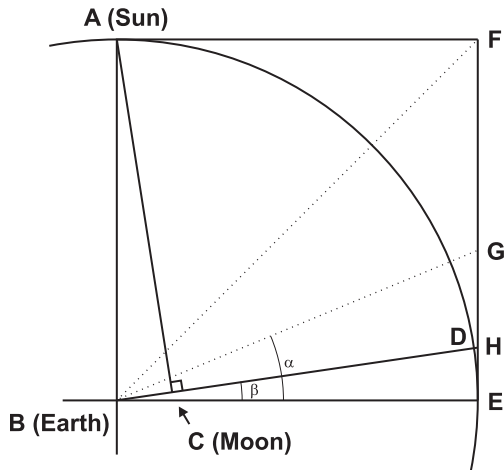



Figure 1.8
Diagram for Proposition 7 of *On the Sizes*

Let *C* be the center of the Moon when halved; let a plane be carried through *AB* and *C*, and let the section made by it in the sphere on which the center of the Sun moves be the great circle *ADE*.

Let *AC*, *CB* be joined, and let *BC* be produced to *D*.

Then, because the point *C* is the center of the Moon when halved, the angle *ACB* will be right.

Let *BE* be drawn from *B* at right angles to *BA*; then the circumference *ED* will be one-thirtieth of the circumference *EDA*; for, by hypothesis, when the Moon appears to us halved, its distance from the sun is less than a quadrant by one-thirtieth of a quadrant.

Thus the angle *EBC* is also one-thirtieth of a right angle.

Let the parallelogram *AE* be completed, and let *BF* be joined.

Then the angle *FBE* will be half a right angle.

Let the angle *FBE* be bisected by the straight line *BG*; therefore the angle *GBE* is one fourth part of a right angle.

But the angle *DBE* is also one thirtieth part of a right angle; therefore the ratio of the angle *GBE* to the angle *DBE* is that which 15 has to 2: for, if a right angle be regarded as divided into 60 equal parts, the angle *GBE* contains 15 of such parts, and the angle *DBE* contains 2.

Now, since *GE* has to *EH* a ratio greater than that which the angle *GBE* has to the angle *DBE*, therefore *GE* has to *EH* a ratio greater than that which 15 has to 2.

Next, since *BE* is equal to *EF*, and the angle at *E* is right, therefore the square on *FB* is double of the square on *BE*.

But, as the square on *FB* is to the square on *BE*, so is the square on *FG* to the square on *GE*; therefore the square on *FG* is double of the square on *GE*.

Now 49 is less than double of 25, so that the square on *FG* has to the square on *GE* a ratio greater than that which 49 has to 25; therefore *FG* also has to *GE* a ratio greater than that which 7 has to 5.

Therefore, *componendo*, *FE* has to *EG* a ratio greater than that which 12 has to 5, that is, that which 36 has to 15.

But it was also proved that *GE* has to *EH* a ratio greater than that which 15 has to 2; therefore, *ex aequali*, *FE* has to *EH* a ratio greater than that which 36 has to 2, that is, than that which 18 has to 1; therefore *FE* is greater than 18 times *EH*.[49]

Explanation: By hypothesis $\angle ABC = 87°$ therefore $\angle DBE = 3°$. Since angles are measured by fractions of right angles, we shall write $\angle ABC = \frac{29}{30}R$ and $\angle DBE = \frac{1}{30}R$, where $R = 90°$.

By construction, $\angle FBE = \frac{1}{2}R$ and $\angle GBE = \frac{1}{4}R$; therefore $\frac{\angle GBE}{\angle DBE} = \frac{15}{2}$.

Since $\frac{GE}{EH} > \frac{\angle GBE}{\angle DBE}$, we know that $\frac{GE}{EH} > \frac{15}{2}$. This key step, which applies the inequality without comment or proof (see below), allows Aristarchus to convert from an equality of *angles* to an inequality of *lengths*.

Now ΔFBE is a right isosceles triangle, so $FB^2 = BE^2$; but (by *Elements* VI.3) $\frac{FB^2}{BE^2} = \frac{FG^2}{GE^2}$, so $FG^2 = 2GE^2$.

Next, since $\sqrt{2} > 7/5$, we have $\frac{FG}{GE} > \frac{7}{5}$, so $\frac{FE}{GE} > 1 + \frac{7}{5} = \frac{12}{5}$.

Finally, $\frac{FE}{EH} = \frac{FE}{GE} \cdot \frac{GE}{EH} > \frac{12}{5} \cdot \frac{15}{2} = 18.$

The lemma applied by Aristarchus in the argument above is stated by Archimedes as follows:

[49] Excerpted from [Heath 1913, 376–381].

> If of two right-angled triangles, (one each of) the sides about the right angle are equal (to each other), while the other sides are unequal, the greater angle of those toward [next to] the unequal sides has to the lesser (angle) a greater ratio than the greater line of those subtending the right angle to the lesser, but a lesser (ratio) than the greater line of those about the right angle to the lesser.[50]

Thus in figure 1.9 where $AC = DF$ and $\alpha > \beta$,

$$\frac{EF}{BC} < \frac{\alpha}{\beta} < \frac{DE}{AB}. \tag{1.1}$$

Aristarchus applies the second inequality to ΔBEH and ΔBEG in figure 1.8; elsewhere in *On the Sizes* he uses both inequalities. We may understand this result in modern terms as follows: since $AC/BC = \sin \alpha$, $DF/EF = \sin \beta$, $AC/AB = \tan \alpha$, and $DF/DE = \tan \beta$,

$$\frac{\sin\alpha}{\sin\beta} < \frac{\alpha}{\beta} < \frac{\tan\alpha}{\tan\beta}.\text{[51]} \tag{1.2}$$

These facts, seen easily from the graphs of the sine and tangent functions, may also be proved geometrically.[52]

We find similar uses of trigonometry in one of the works of deservedly the most famous of ancient scientists, Archimedes (287–212 BC). One of the most creative mathematicians of all time, Archimedes is renowned for his ability to approach difficult geometric problems, including his determinations of areas of curved figures. His “method of exhaustion” works by gradually inscribing, and simultaneously circumscribing, the required figure with polygons that successively approximate the curves more closely. Archimedes also wrote a number of works in topics that would now be classified under physics, including mechanics, optics, and hydrostatics. His feared war machines and other instruments spread his fame far beyond the confines of science.[53]

[50] From the *Sand Reckoner*, translated by Wilbur Knorr in [Knorr 1985]. The origin of this lemma is lost, but it has been speculated that it was part of a pre-Euclidean collection of lemmas for use in the study of spherics [Hultsch 1883].

[51] See, for instance, [Heath 1921, vol. 2, 5], and similar statements in [Zeller 1944], 3–4.

[52] One such proof is given in [Heath 1913, 366].

[53] See [Clagett 1970] for an extensive biographical essay on Archimedes. The two leading modern sources for Archimedes are the translations by Thomas Little Heath in [Archimedes 1897] (including most of the *Sand Reckoner*) and the accounts of [Dijksterhuis 1956], especially the Princeton 1987 reprint which contains an essay by Wilbur Knorr bringing the state of Archimedes scholarship up to date.

Figure 1.9
An ancient trigonometric lemma

ADMIRANDI
ARCHIMEDIS
SYRACVSANI
MONVMENTA OMNIA MATHEMATICA, QVAE EXTANT,
Quorumque Catalogum inuersa Pagina demonstrat,
EX TRADITIONE DOCTISSIMI VIRI
D. FRANCISCI MAVROLICI
NOBILIS SICVLI, ABBATIS SANCTAE MARIAE A PARTV.
OPVS
Præclarissimum, non priùs Typis commissum, à Mathefeos verò Studiosis enixè desideratum, tandemque è fuligine Temporum accuratè excussum.
AD ILLVST. ET RELIGIOSISSIMVM VIRVM
FR. SIMONEM
RONDINELLI
SAC. HIEROSOLYMITANAE RELIGIONIS EQVITEM LAVDATISS.
S. Ioannis Baptistæ à Sauigliana, necnon Pondaderæ, & S. Philippi de Osmo Commēdatarium digniss. Vnius è Melitensibus Triremibus olim strenuissimum Ductorē, plurimarumq; Nauium Turcicarum Debellatorē gloriosum, In Vrbe feliciss. Panormo pro sua Sac. Relig. pluribus annis vigilantiss. Legatum, Receptorem, ac Procuratorem Generalem, & inclytæ Reaccensorum Academiæ Vrbe in ipsa eruditissimum Principem: &c.
PANORMI
Apud D. Cyllenium Hesperium, Cum Licentia Superiorum.
MDC. LXXXV.
SVMPT. ANTONINI GIARDINAE, BIBLIOPOLAE PANORM.

Figure 1.10
Title page of Maurolico's 1685 edition of Archimedes' collected works, including the *Sand Reckoner* (courtesy of the Burndy Library)

The *Sand Reckoner* (see figure 1.10), one of several of Archimedes's works intended for a wider audience, was written to show the difference between a large finite quantity and the infinite. He illustrates this amply by calculating the number of grains of sand that would be required to fill the universe. His result, an upper bound on the order of 10^{63}, was astronomically large

(and required the invention of an ingenious number system to express), but not infinite.[54] To make sure that he had truly capped the number with an upper bound Archimedes needed to use the largest available estimate for the size of the universe, hence his use of Aristarchus's heliocentric system. If the Earth travels around the Sun, then the vantage point from which we see the stars is moving, so they should appear to move over the course of the year. In fact they do, but by such a tiny amount that the effect is nowhere near visible. Thus Aristarchus had to move the fixed stars far enough away that this ***stellar parallax*** was negligible, resulting in the largest universe known to Archimedes.

The *Sand Reckoner* begins with a measurement of the angular magnitude of the diameter of the Sun, d_s, which Archimedes reports to be

$$\frac{1}{200}R < d_s < \frac{1}{164}R \tag{1.3}$$

(where $R = 90°$). (It had been thought that Archimedes obtained these bounds by drawing lines of sight on his observational instrument and measuring their angles physically. Recently, a procedure has been reconstructed that uses mathematical methods similar to what follows, including an application of (1.1).[55]) Using (1.3), Archimedes needs to find a lower bound on the length of the Sun's diameter. Since he already has Aristarchus's estimate of the size of the circle in which the Sun orbits, Archimedes needs only to find a bound relative to that circle.

In figure 1.11 (see also figure 1.12), E is the observer at a moment after the Sun has risen, $\angle PEQ$ is the Sun's angular diameter seen from E, and $\angle FCG$ its angular diameter seen from the center of the Earth C. A and B are the extensions of the lines of sight from C. From (1.3) we know that $\angle PEQ < \frac{1}{164}R$. Since E is closer to the Sun than C, the angular diameter seen from C is smaller, so $\angle FCG < \frac{1}{164}R$. Thus, the arc $\widehat{AOB}$ in the corresponding circle is less than $\frac{1}{164 \times 4} = \frac{1}{656}$ of a circle, and AB is less than one side of a 656-gon. Hence $AB / CO < \frac{1}{656} \cdot 2\pi < \frac{1}{100}$.[56] But AB is the Sun's diameter,[57] and since the Earth is assumed to be smaller than the Sun, the sum of the Earth's and Sun's radii is less than $\frac{1}{100}CO$. Therefore $\frac{100}{99} > CO / HK > CF / EQ$.[58]

[54] For an extended but accessible discussion of Archimedes' purpose in this work, see [Netz 2003].

[55] See [Delambre 1817, vol. 1, 104–105], [Hultsch 1899, 197], and [Lejeune 1947, 38–41] for the original theories, and [Shapiro 1975] for his rebuttal and reconstruction of Archimedes' method.

[56] Here Archimedes applies his lower bound for π, 22/7, from *Measurement of a Circle*.

[57] Since $CA = CO$, AM is perpendicular to CO, and OF is perpendicular to AC, by symmetry we have $AM = OF$, and the latter is the Sun's radius.

[58] The first of these inequalities is true because HK is what is left over when the two radii are taken away from CO. The second inequality is true because $CO > CF$ and $HK < EQ$.

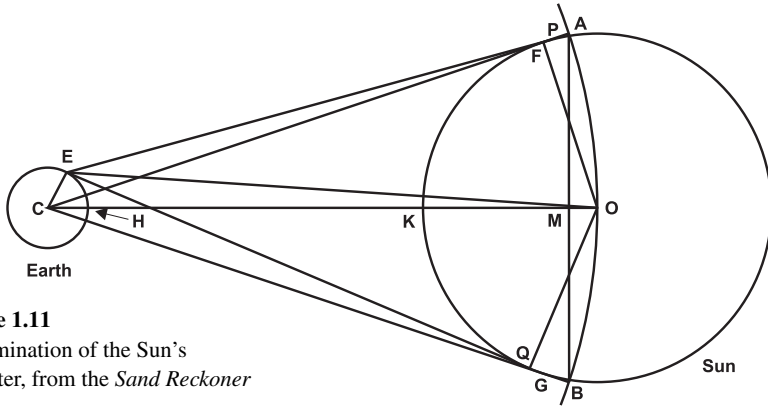



Figure 1.11
Determination of the Sun's diameter, from the *Sand Reckoner*

DE ARENAE NVMERO. 52

centum ad nouem & nonaginta. angulus igitur l d x ad m h o angulum minorem proportionem habet, quàm centum ad nouem & nonaginta. & quoniam angulus l d x maior eſt, quàm ducenteſima pars anguli recti: erit angulus m h o maior, quàm nouem & nonaginta partes anguli recti in uiginti millia partium diuiſi. quare maior, quàm una pars anguli recti diuiſi in ducentas & tres partes. linea ergo b a maior eſt, quàm quæ ſubtenditur uni parti circumferentiæ circuli a b c, diuiſæ in partes octingentas & duodecim. ſed ipſi b a ſolis diameter eſt æqualis. manifeſtum eſt igitur diametrum ſolis maiorem eſſe latere figuræ, quæ mille angulis conſtet. Itaque his poſitis, oſtenduntur & illa: uidelicet diametrum mundi continere minus, quàm decies millies diametrum terræ: & inſuper diametrum mundi minorem eſſe, quàm centum myriadum myriadum ſtadiorum. Quoniam enim poſitum eſt, diametrum ſolis non maiorem eſſe, quàm trigintuplam diametri lunæ: diametrum uero terræ diametro lunæ maiorem: conſtat diametrum ſolis minorem eſſe, quàm trigintuplam diametri terræ. Rurſus quoniam oſtenſum eſt diametrum ſolis maiorem eſſe latere figuræ mille angulorum, quæ in maximo mundi

Figure 1.12
The equivalent to figure 1.11 in Commandino's edition of Archimedes' *Sand Reckoner* (courtesy of the Burndy Library)

Now Archimedes applies (1.1) to ΔOEQ and ΔOCF (since $OF = OQ$ but $EQ < CF$, since E is closer to the Sun than C), which gives

$$\frac{CO}{EO} < \frac{\angle OEQ}{\angle OCF} < \frac{CF}{EQ}.$$

This result allows him to convert from his inequality of lengths to one of angles. Doubling both angles and recalling that we already know that $\angle PEQ > \frac{1}{200}R$, it is relatively easy sailing (using the second inequality above) to arrive at $\angle ACB > \frac{99}{20000}R > \frac{1}{203}R$, so that $\widehat{AB}$ is greater than $\frac{1}{203\times 4} = \frac{1}{812}$ of a circle, and finally that AB is greater than one side of an 812-gon. Archimedes rounds downward to the side of a chiliagon (1000-gon), and has a convenient lower bound for the Sun's diameter.

That both Aristarchus and Archimedes use (1.1) with no comment or proof suggests that the lemma would have been familiar to their readers. In fact there are nine proofs of one or the other inequality in this lemma in antiquity, but the earliest proof of the first inequality does not appear until Ptolemy's construction of his chord table in the *Almagest* four centuries later.[59] How the lemma might have arisen may only be speculated. Perhaps it was recorded in a mathematical work alongside and adjunct to the study of spherics,[60] or perhaps it was more closely bound with the emergence of practical astronomical problems.[61] The lemma is useful, for instance, in the determination of rising times of arcs of the ecliptic, or in the measurement of time with sundials.[62]

Whether or not the lemma allows us to say that a more extensive trigonometry existed as early as the third century BC is a matter of debate.[63] Clearly its existence alone does not qualify as sufficient evidence. Nor does it provide us with a useful general trigonometric tool; although it is perfectly plausible to measure angles as fractions of a right angle, the lemma gives reasonable and useful bounds only for very small arcs.

[59] [Knorr 1985] gives a detailed comparison of these nine sources, noting their remarkable similarities. An equivalent to the tangent half of the lemma appears in Euclid's *Optics*, in a different context and format. Opinions on which is the earliest text to contain this proof have varied; see [Knorr 1991, 195], [Knorr 1994], and [Jones 1994].

[60] As suggested in [Hultsch 1883], especially pp. 415–420.

[61] This view is favored by [Björnbo 1902].

[62] [Knorr 1985, 383–384].

[63] Views on this matter go back at least to the nineteenth century; see for instance [Tannery 1893, 60–68]. More recently, Toomer suggests that lemma (1.1) may have been "of fundamental importance in evaluating triangles before the development of trigonometry" [Diocles 1976, 162], a view that Knorr is inclined to support in [Knorr 1985, 385].

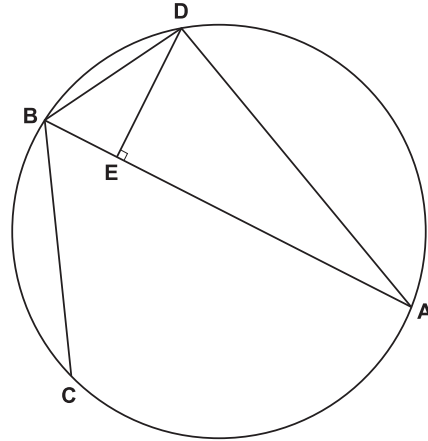



Figure 1.13
Archimedes' Theorem of the Broken Chord

Given an arbitrary arc of sufficient magnitude, the lemma cannot produce bounds of any use to an astronomer in a reasonable time. On its own, then, (1.1) is unable to provide a systematic conversion between arcs and lengths.

We are not yet finished with Archimedes. In his treatise *Book on the Derivation of Chords in a Circle*, the early eleventh-century Muslim scientist al-Bīrūnī attributed the following result to Archimedes, sometimes called the "Theorem of the Broken Chord": in figure 1.13, given unequal arcs $\widehat{AB} > \widehat{BC}$ in a circle, let D be the midpoint of $\widehat{AC}$ and drop perpendicular DE onto AB. Then $AE = EB + BC$.[64] From here it is possible to derive results equivalent to the sine or chord angle sum and difference identities, among others. This fact led one scholar to claim that Archimedes had a more sophisticated trigonometry than what we have witnessed thus far, one that would have given him the capacity to construct a table of chords as powerful as Ptolemy's.[65] However, the trigonometric identities do not drop entirely without effort from this theorem,[66]

[64] [Al-Bīrūnī 1948, 7–8], a German translation along with commentary and the appearance of the same theorem in works of other medieval Arabic authors is available in [Suter 1910–11]. Al-Bīrūnī also mentions this theorem in his *Masᶜudic Canon* [Al-Bīrūnī 1954, 273–274]; the relevant section is translated into German in [Schoy 1927, 2–7, esp. 3–4]. See also the exposition of al-Bīrūnī's text in [Dold-Samplonius 1973, 34–39].

[65] [Tropfke 1928]; announced in [Miller 1928] and repeated elsewhere; see for instance [Karpinski 1945, 269].

[66] Toomer derives the chord difference formula, for instance, as follows: let O be the center of the circle, F the point diametrically opposite from D, $\alpha = \angle AOD$, and $\beta = \angle DOB$. Now, $\Delta DEB \sim \Delta DFA$ (since $\angle DBE = \angle AFD$ subtend the same chord DA—thanks to Alina Kononov and Eric Gorlin for this) and $\Delta AED \sim \Delta FBD$ ($\angle BAD = \angle BFD$ since they subtend the same chord BD). So $BE = (BD \cdot AF) / DF$ and $AE = (AD \cdot BF) / DF$. Then, from Archimedes' theorem,

and, just as seriously, we have no idea how Archimedes (as opposed to his Muslim successors) might have intended to use it.[67] It can only be said that, had Archimedes the idea and numerical system needed to build a systematic chord table, he would have had the mathematics at his disposal to pull it off.[68]

$BC = \text{Crd}(\alpha - \beta) = AE - BE = (AD \cdot BF - BD \cdot AF) / DF$. Since $AD = \text{Crd } \alpha$, $BD = \text{Crd } \beta$, DF is the diameter, and BF and AF may be found by applying the Pythagorean Theorem with known quantities, the chord difference law results.

[67] This argument is made convincingly in [Toomer 1973a, 20–23].

[68] See [Schneider 1979, 149–151] for an argument to this effect. Other cases have been made for more extensive calculations with chords in the third century BC than are found in the extant literature; see for instance [Tannery 1893, 60–68], [van der Waerden 1970, 7], and especially [van der Waerden 1986, esp. 400–401] for their views on Apollonius, and [Szabó 1985a] on Eudoxus. Since these arguments must remain speculative (barring further discoveries), we simply refer the interested reader to these articles. With respect to Archimedes, we have also his *Measurement of a Circle*. In his computation of bounds for the value of π he uses the equivalent of a chord half-angle formula, which we shall see is a vital tool for the construction of a chord table. Once again, we have no evidence that Archimedes used it outside of the context of computing π.